440-TP-012-001

NASA'S MISSION TO PLANET EARTH

EARTH PROBES

DATA INFORMATION SYSTEM

EOS

EARTH OBSERVING SYSTEM
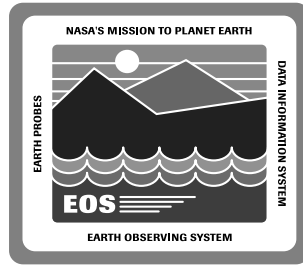
Planning Subsystem Database Size Estimate for the ECS Project

Technical Paper—Not intended for formal review or government approval.

Technical Paper

February 1995

RESPONSIBLE ENGINEER

Mac McDonald for Chhaya Rao	2/17/95

Chhaya S. Rao, PDPS Scientist/Engineer   Date
EOSDIS Core System Project

SUBMITTED BY

Parag N. Ambardekar	2/17/95

Parag Ambardekar, PDPS Manager   Date
EOSDIS Core System Project

Hughes Applied Information Systems
Landover, Maryland

This page intentionally left blank.

Abstract

The purpose of this paper is to define a model for the Planning Subsystem database and then determine its size and the number of database accesses it would have to support. This paper is a preliminary attempt in deriving estimates based on the current understanding of the Planning Subsystem design. The gross estimates are based on a rough model of the Planning database and then determining what the size would be to support one month's worth of processing. CERES and MODIS workloads based on the Ad Hoc Working Group on Production data for the Releases A (epoch c) and B (epoch F) are analyzed and the database size estimated. These calculations are gross estimations and will be revised during the CDR timeframe.

Keywords: Planning Subsystem, Planning database, PGE, AHWGP, PDPS, DBMS, CERES, MODIS

Abstract

1. Introduction

2. Planning Database Data Dictionary

3. Planning Database Record Size Estimates

4. Planning (CERES) Database Size Estimates (Epoch c)

5. Planning (CERES) Database Size Estimates (Epoch F)

6. Planning (MODIS) Database Size Estimates (Epoch F)

Abbreviations and Acronyms

Tables

This page intentionally left blank.

1. Introduction

1.1 Purpose and Scope

The purpose of this paper is to define a model for the Planning Subsystem database, and then determine its size and the number of database accesses it would have to support.

The estimates presented in this paper are derived gross estimates based on the current understanding of the Planning subsystem design. They will be further refined as the system design matures. The Planning database size was estimated by first developing a rough model of the Planning subsystem database and then determining what its size would be to support one month's worth of CERES processing during the EPOCH c timeframe. The number of daily Planning database accesses was calculated by first partitioning the Planning subsystem's daily activities into 5 major functions. For each of these functions the number of database accesses required to support one day's worth of planning during Epoch c was estimated. (This estimate represents the amount of daily data traffic the Planning database would have to support.) This exercise was repeated for CERES and MODIS workloads during Epoch F. The Planning database model, the collected data, the calculations and the drawn conclusions have been presented in this paper.

The Planning database has the following major object classes:

- PGE Profile
- Production Requests
- Data Processing Requests
- Plans
- Resource Information
- Data Availability Schedules
- Planning Subsystem Parameters

On a typical day, the Planning subsystem would perform the following major functions:

- Production Request processing.
- Plan creation
- Plan activation
- Plan feedback updates
- Report generation

1.2   Organization

The paper is organized as follows:

- Section 1 provides the purpose and scope of the paper, its organization and review and approval procedures and executive summary.

- Section 2 defines a model of the Planning database conceptual data dictionary.

- Section 3 presents estimates for the Planning (CERES) database during EPOCH c - database size for one month and number of database accesses per day.

- Section 4 presents estimates for the Planning (CERES) database during EPOCH F - database size for one month and number of database accesses per day.

- Section 5 presents estimates for the Planning (MODIS) database during EPOCH F - database size for one month and number of database accesses per day.

- The Appendices provide information on the abbreviations and acronyms used.

1.3   Acknowledgments

1.4   Review and Approval

Questions regarding technical information contained within this paper should be addressed to the following ECS and/or GSFC contacts:

- ECS Contact:

  Mac McDonald, PDPS System Engineer

  (301)-925-0364

  mac@eos.hitc.com

- GSFC Contact:

  Steve Kempler, PDPS Manager

  (301)-286-7766

  steven.j.kempler@gsfc.nasa.gov

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems
1616 McCormick Dr.
Landover, MD 20785

1.5   Applicable and Reference Documents

1. ECS Core Metadata Baseline; October 1994

1.6 Executive Summary

The Planning database was sized for CERES data during Epoch c and Epoch F. It assumed one month's worth of data. Additional assumptions are given in Sections 3.1 and 4.1 resp.

The Planning database size for CERES data:

Epoch c: 164 Mbytes

Epoch F: 655 Mbytes

The number of read/ write accesses (of 12 Kbytes each) into the Planning CERES database per day:

Epoch c : 50,400

Epoch F : 200,200

During Epoch c, background load for CERES processing: 50,400 / 16= 3150 accesses per hour = 3150/60 = 52.5 accesses per min or less than 1 access per second.

During Epoch c, peak load (plan creation/activation): upto 30,000 accesses/plan. Assuming a plan takes 5 minutes to generate, this is equivalent to about 100 accesses/sec.

During Epoch F, background load for CERES processing: 200,800 / 16= 12,550 accesses per hour = 12,550/60 = 209.1 accesses per min or less than 4 accesses per second.

During Epoch F, peak load (plan creation/activation): upto 120,000 accesses/plan. Assuming a plan takes 8 - 10 minutes to generate, this is equivalent to about 200 accesses/sec.

The Planning database was sized for MODIS data during Epoch F. It assumed one month's worth of data. Additional assumptions are given in Sections 5.1 .

The Planning database size for MODIS data:

Epoch F : 9.5 Gbytes

The number of read/ write accesses (of 12 Kbytes each) into the Planning MODIS database per day:

Epoch F : 3,190,000

During Epoch F, background load for MODIS processing: 3,190,000 / 16= 199,375 accesses per hour = 199,375 / 60 = 3323 accesses per min or less than 55 accesses per second.

During Epoch F, peak load (plan creation/activation): upto 1,880,000 accesses/plan. Assuming a plan takes 10- 12 minutes to generate, this is equivalent to about 2600 accesses/sec.

This page intentionally left blank.

2. Planning Database Data Dictionary

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (1 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
PGE Profile	AI&T	PGE (Algorithm) Identifier	Unique identifier for each PGE, needed to generate specific products.	char(8)		
PGE Profile	AI&T	PGE (Algorithm) Package Name	This attribute is the name given to the complete delivered PGE package submitted for algorithm integration and test	char (80)		
PGE Profile	AI&T	PGE (Algorithm) Package Characteristics	This attribute specifies the data set, version, series and instance of the full PGE package delivered. Also indicates if this is the default PGE.	char(20)		
PGE Profile	AI&T	PGE Priority	This attribute specifies the execution priority of the PGE.	char(1)		
PGE Profile	AI&T	PGE Input Data Dependency Profile	This attribute identifies the input data needed, conditions for use, size of data , quality check flags, and the data server (s) that contains this data.	char(50) per input file - or - char(10,000) for 200 input files per PGE		
PGE Profile	AI&T	PGE Output Data Identifier	This attribute identifies the type of output data product to be generated.	char (80) per output file		
PGE Profile	AI&T	PGE Output Data Profile	This attribute identifies the output data product characteristics, size of data and the data server(s) that will store this data.	-or- char(800) for 10 output files		
PGE Profile	AI&T	PGE Resource Profile	This attribute identifies the resources needed to support execution of the PGE, such as suitable supporting hardware platforms, O.S. required, CPU consumption, expected run times, and storage requirements.	char (100)		

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (2 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
PGE Profile	AI&T	PGE Process Control Parameters	This attribute identifies the data containing PGE process control parameters and PGE dependencies that need to be considered during planning. For example: process parameters in the processing metadata (ex.: instrument mode indicator, etc.), quality flag indicators, implementation alternatives and conditions (ex.: timeout rules), temporal coverage overlap, tessellation (mosaicing / spatial overlap), etc.	char(200)		
PGE Profile	AI&T	PGE Attention Flags	This attribute indicates special aspects of each PGE such as Test mode execution only flags, QA check flags, conditions for not running PGE, need for existence of prior instance of product, or existence of a certain number of previous instances of products, etc.	char(15)		
PGE Profile	AI&T	PGE Product Chain Identifiers	This attribute specifies the product chain(s) the PGE is associated with.	char(150)		
PGE Profile	AI&T	PGE Activation Rules	This attribute identifies the conditions under which a PGE is planned and subsequently scheduled. This supports conditional activation based on metacharacteristics needed to execute a PGE (predictable: day/night only indicator; unpredictable: observation condition (cloud cover, snow), data quality (missing data, saturation data), science context (volcano, flood), etc.).	char(200)		
PGE Profile	AI&T	PGE Process Script Template Files	This attribute specifies the "recipes" and control parameters (details of paths to be traversed) needed for the execution of each PGE executable by identifying the process execution script template filenames .	char(20) for each template -or- char(200) for 10 templates		

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (3 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
PGE Profile	AI&T / Operations Staff	PGE Profile Last Update Date	This attribute specifies the date when the PGE Profile was last updated.	yy/dd/hh mm:ss:tt char(12)		
Production Requests	Operations Staff / Science user	Production Request Identifier	This attribute specifies a unique ID to identify the production request. ID indicates whether request is a standard/reprocessing/on-demand request.	char(8)		
Production Requests	Operations Staff / Science user	Production Request User Information	This attribute specifies information about the user submitting the request such as authorized user ID, contact address, etc.	char(50)		
Production Requests	Planning CSCI	Production Request Priority	This attribute specifies the priority assigned to the production request.	char(3)		
Production Requests	Operations Staff / Science user	PGE Output Data Identifier	This attribute identifies the type of output data product to be generated.	char(30)		
Production Requests	Operations Staff / Science user	PGE (Algorithm) Identifier	This attribute specifies the identifier of the PGE to be executed.	char(8)		
Production Requests	Operations Staff / Science user	PGE (Algorithm) Package Characteristics	This attribute specifies the data set, version, series and instance of the PGE to be executed.	char(80)		
Production Requests	Operations Staff / Science user	Production Request Input Data	This attribute identifies the input data to be processed such as time windows of receipt of specific data sets / geographic locations.	char(50)		
Production Requests	Operations Staff / Science user	Production Request Output Data Characteristics	This attribute specifies the output data product characteristics to be generated such as output data destination.	char(50)		

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (4 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
Production Requests	Operations Staff / Science user	Production Request User-Specified Parameters	This attribute includes user specified special processing instructions such as time deadlines (latest planned completion time), priorities alternate inputs, and parameters relevant to the production request.	char(80)		
Production Requests	Operations Staff / Science user	Production Request Date/ Time Entered	This attribute specifies the date / time production request had been entered.	yy/dd/hh: mm/ss/tt or char (12)		
Production Requests	Planning CSCI	Production Request Date/ Time Closed	This attribute specifies the date / time production request had been closed.	yy/dd/hh: mm/ss/tt or char (12)		
Production Requests	Planning CSCI	Production Request Status information	This attribute specifies the current production request status (in progress, closed, canceled, inactive) and date status was entered.	char(3), yy/dd/hh: mm/ss/tt or char (15)		
Production Requests	Planning CSCI	Production Request -Associated Data Processing Requests	This attribute specifies all the Data Processing Requests associated with the Production Request.	char(8) or char(1200) for 150 DPRs per production request.		
Production Requests	Operations Staff / Science user	Production Request Last Update Date	This attribute specifies the date when the Production Request was last updated.	yy/dd/hh mm:ss:tt or char (12)		
Data Processing Requests	Planning CSCI	Data Processing Request Identifier	This attribute identifies the Data Processing Request ID.	char(8)		
Data Processing Requests	Planning CSCI	Production Request Identifier	This attribute specifies the production request identifier associated with the DPR.	char(8)		
Data Processing Requests	Planning CSCI	PGE (Algorithm) Identifier	This attribute specifies the ID of the PGE to be executed.	char(8)		

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (5 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
Data Processing Requests	Planning CSCI	PGE (Algorithm) Package Characteristics	This attribute specifies the data set, version, series and instance of the PGE to be executed.	char(80)		
Data Processing Requests	Planning CSCI	Data Processing Request Priority History	This attribute tracks the priority assigned to the DPR and date it was assigned.	char(20)		
Data Processing Requests	Planning CSCI	PGE Input Granule Identifiers / Availability / Quality	This attribute specifies the input data granule identifiers (URs) that include data set, version, series and instance of input data information. In addition, it specifies the availability of the input data and identifies alternative input data granules. It also indicates whether the quality of the input data needs to be checked.	char(75) per input file or char(15,000) for 200 input files per DPR..		
Data Processing Requests	Planning CSCI	PGE Input Granule Characteristics	This attribute specifies the input data characteristics such as input data size and the data server(s) storing this information. It also specifies input data metacharacteristics such as sensor mode, etc. if applicable.			
Data Processing Requests	Planning CSCI	PGE Output Data Identifier	This attribute identifies the type of output data product to be generated.	char(100) per output file or char(1000) for 10 output files		
Data Processing Requests	Planning CSCI	PGE Output Data Characteristics	This attribute specifies the desired output data / metadata UR, any significant production history related data, size of data and data server ID(s) for storage.			

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (6 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
Data Processing Requests	Planning CSCI	PGE Resource Profile	This attribute identifies the resources needed to support execution of the PGE, such as suitable hardware platforms, O.S. required, CPU consumption, and storage requirements.	char(100)		
Data Processing Requests	Planning CSCI	Data Processing request-specific parameters	This attribute includes special processing parameters specific to this DPR such as processing time deadlines, special user specified instructions / parameters, Test mode execution only flags, QA check flags, conditions for not running PGE, need for existence of prior instance of product, or existence of a certain number of previous instances of products, etc.	char(100)		
Data Processing Requests	Planning CSCI	PGE Planned Execution Start Time	This attribute specifies the predicted execution start time.	yy/dd/hh mm:ss:tt or char(12)		
Data Processing Requests	Planning CSCI	PGE Planned Execution Stop Time	This attribute specifies the predicted execution stop time.	yy/dd/hh mm:ss:tt or char(12)		
Data Processing Requests	Planning CSCI	PGE Actual Execution Start Time	This attribute specifies the actual execution start time.	yy/dd/hh mm:ss:tt or char(12)		
Data Processing Requests	Planning CSCI	PGE Actual Execution Stop Time	This attribute specifies the actual execution stop time.	yy/dd/hh mm:ss:tt or char(12)		
Data Processing Requests	Planning CSCI	Data Processing Request Descendent Trace	This attribute specifies all descendent data processing requests that are dependent on this data processing request.	char(40)		

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (7 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
Data Processing Requests	Planning CSCI	Data Processing Request Status History	This attribute tracks the status of the data processing request and date status is entered. It also specifies the reason for any failure in execution such as bad/missing data, hardware failure, etc.	char(3) yy/dd/hh mm:ss:tt or char(100) to track 6 statuses		
Data Processing Requests	Planning CSCI	Data Processing Request Production History	This attribute specifies historical Production history data such as Quality flag indicators, and any changes to the processing specifications entered in the Data Processing Request such as a change in resources, CPU utilization, alternate input data used, reasons for delay in processing, changes entered by the Processing operator, descendent DPRs canceled due to this DPR, etc.	char(1000) per output file or char(10,000) for 10 output files		
Data Processing Requests	Planning CSCI	PGE Process Script Template Files	This attribute specifies the "recipes" and control parameters (details of paths to be traversed) needed for the execution of each PGE executable by identifying the process execution script template filenames. Exception handling template files are also identified.	char(20) per template or char(200) for 10 templates		
Plans	Planning CI	Plan Status	This attribute specifies whether the plan is an active plan or a candidate plan.	char(3)		
Plans	Planning CI	Plan Identifier	This attribute specifies the Plan ID. Identifier includes the DAAC ID.	char(8)		
Plans	Planning CI	Plan Creation Time	This attribute specifies the plan creation time.	yy/dd/hh mm:ss:tt or char(12)		
Plans	Planning CI	Plan Start Time	This attribute specifies the plan start time.	yy/dd/hh mm:ss:tt or char(12)		

440-TP-012-001

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (8 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
Plans	Planning CI	Plan Stop Time	This attribute specifies the plan stop time.	yy/dd/hh mm:ss:tt or char(12)		
Plans	Planning CI	Data Processing Request Identifier	This attribute identifies the Data Processing Request ID.	char(8) or 12,000 for 1500 DPRs		
Plans	Planning CI	Plan Generation Environment	This attribute specifies where the database environment used to generate the plan is stored.	char(100)		
Production Rules	Planning CI	DAAC-specific Production Request Priorities	This attribute specifies the valid Production Request Types and the associated priorities.	char(10)		
Production Rules	Planning CI	DAAC-specific Production Request Workload	This attribute specifies the % of processing workload assigned to processing standard/ reprocessing / on-demand production requests.	char(10)		
Production Rules	Planning CI	DAAC-specific Instrument Processing Priorities	This attribute specifies the instrument processing priorities relative to each other.	char(10)		
Production Rules	Planning CI	DAAC-specific Product Priority	This attribute specifies the priority of one product over another.	char(10)		
Production Rules	Planning CI	DAAC-specific Instrument Resource Constraints	This attribute specifies any resource constraints concerning the processing of specified instrument data.	char(10)		
Production Rules	Planning CI	DAAC-specific First Choice Hardware	This attribute specifies the first choice hardware based on (a) Production Request Type (b) Product type.	char(10)		
Production Rules	Planning CI	Production Rules Last Update Date	This attribute specifies the date when the Production Rules were last updated.	yy/dd/hh mm:ss:tt or char(12)		

440-TP-012-001

Table 2-1. Planning CSCI Database Conceptual Data Dictionary (9 of 10)

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
Predicted Resource Availability	CSMS / MSS	Site Identification	This attribute identifies the site where the resource is located.	char(6)		
Predicted Resource Availability	CSMS / MSS	Site-Specific Resource Identification	This attribute identifies the resource at a specific site.	char(15)		
Predicted Resource Availability	CSMS / MSS	Resource characteristics	This attribute specifies the characteristics associated with a given resource (such as processing capacity (MFLOPS) and RAM for processing platforms, O.S. supported, storage capacity for storage devices, etc.	char(100)		
Predicted Resource Availability	CSMS / MSS	Resource Availability Status	This attribute indicates whether the resource is available for allocation and date status was updated.	char(3)		
Predicted Resource Availability	Planning CI	Resource Usage	This attribute specifies any usage of the resource already planned for a specific time frame and unused capacity.	char(10)		
Data Availability Schedules	Planning CI	Data Availability DAAC Site	This attribute specifies the DAAC site at which the data is available.	char(6)		
Data Availability Schedules	Planning CI	Data Instrument Type	This attribute specifies the instrument(s) associated with the data.	char(6)		
Data Availability Schedules	Planning CI	Data Time Window of Receipt	This attribute specifies the time window of receipt of the instrument data.	char(24)		
Data Availability Schedules	Planning CI	Data Identification	This attribute identifies the UR of the data granule / product that is available.	char(15)		
Data Availability Schedules	Planning CI	Data Availability Date	This attribute specifies the expected availability date of the data.	char(12)		

Object Class	Source	Attribute	Attribute Description	Data Type	Domain Value	Domain Description
Data Availability Schedules	Planning CI	Data Availability Notification Date	This attribute indicates the date notice was received concerning availability of the data.	char(12)		
Data Availability Schedules	Planning CI	Data Location	This attribute specifies the data server location where this data is stored.	char(15)		
Planning Subsystem Parameters	Planning CI	Time Delay Thresholds	This attribute specifies the pre-defined (adapted) data processing request time delay thresholds after which operator should be alerted.	char(20)		
Planning Subsystem Parameters	Planning CI	Plan Deviation Thresholds	This attribute specifies the pre-defined (adapted) plan deviation thresholds. If actual processing deviates from original plan by these thresholds, then the operator should be alerted,	char(20)		
Planning Subsystem Parameters	Planning CI	On-Demand Production Request Acceptance Criteria	This attribute specifies the criteria against which On-Demand production requests are validated such as valid requester IDs, requester ranking / override capability, etc.	char(50)		
Planning Subsystem Parameters	Planning CI	DAAC Sites	This attribute specifies the valid DAAC sites supported.	char(40)		
Planning Subsystem Parameters	Planning CI	Local DAAC	This attribute specifies the DAAC site of the database.	char(5)		

3. Planning Database Record Size Estimates

3.1 Planning Database Record Size Estimates

The following provides the Planning database record size estimates. Some of the Planing records have a static size whereas the size of some of the other records are a function of one or more of the following variables:

- Average number of data processing requests per production request = dpr.

- Average number of input files per data processing request = inputf.

- Average number of ouput files per data processing request = outputf.

- Average number of templates per PGE: tplates.

1. PGE Profile record size (one per PGE): 1000 bytes (approx. base) + 50 bytes X (inputf) + 80 bytes X (outputf) + 20 bytes X (tplates).

2. Production Request record size (one per production request): 400 bytes (approx. base) + 8 bytes X (dpr).

3. Data Processing Request record size (one per data processing request): 800 bytes (approx. base) + 75 bytes X (inputf) + 1100 bytes X (outputf) + 20 bytes X (tplates).

4. Plan record size (one per plan): 147 bytes + 8 bytes X (dpr)

5. Production Rules area: 100 bytes (static)

6. Resource Availability record size (one per resource): 134 bytes

7. Data Availability Schedule record size (one per product generated): 90 bytes

8. Planning Subsystem Parameters: 100 bytes (approx.)

This page intentionally left blank.

4. Planning (CERES) Database Size Estimates (Epoch c)

4.1 Assumptions

The following has been estimated for CERES data using Epoch c average numbers:

1. Estimated for 1 month's worth of data.

2. 8 CERES PGEs supported

3. Only Standard Production requests considered.

Average number of Standard Production Requests per month: 8 - 10.

(To include reprocessing requests (multiply by 4): average number of production requests: 40 / month).

4. Average number of CERES Data Processing Request (corresponding to standard production requests) per day: 50 (based on approx. 50 activations per day). or 1500 Data Processing Requests per month (30 days).

(Average number of CERES DPRs per day corresponding to production requests in general: 200 or 6,000 per month).

5. Average number of input files per CERES PGE: 200

6. Average number of output files per CERES PGE: 10

7. Assume about 150 Data Processing Requests per Production Request

8. Assume about 30 candidate plans and 30 active plans generated per month.

9. Assume about 10 process script templates associated with each PGE.

10. Assume database tracks only process script template filenames and actual templates stored on local disk storage.

11. Assume database environment used to generate each plan tracked is stored on local disk and the Planning database tracks only its location.

12. Assume 14 CERES resources are tracked by the Planning database.

13. Assume a fudge factor (of 2.69) not included when calculating database size.

14. Assume the Planning database tracks the following major data classes:

    - PGE Profile
    - Production Requests
    - Data Processing Requests
    - Plans

- Production Rules

- Predicted Resource Availability

- Data Availability Schedules

- Planning Subsystem Parameters

15. Estimation Rationale: Number of estimated accesses into the database is dependent on the number of processing activations per day per CERES instrument at LaRC.

4.2   Size Estimates of Planning (CERES) Database - Epoch c

One month's worth using EPOCH c statistics

(1) PGE Profile: 12 Kbytes per record

For 8 - 10 PGE records: 120 Kbytes (approx.) - Static value

(2) Production Requests: 1600 bytes

For 10 Standard Production Requests per month: 16 Kbytes

(For 40 Standard / Reprocessing Production Requests per month: 64 Kbytes)

(3) Data Processing Requests: 27 Kbytes per DPR

For 1500 Data Processing Requests: (approx. 40 Mbytes) per month

(For 6000 Data Processing Requests (approx. 160 Mbytes) per month

(4) Plan: 12 Kbytes per plan (for 1500 standard DPRs per plan or 48 Kbytes per plan for 6000 DPRs per month)

For 60 plans per month: 2880 Kbytes

(5) Production Rules: 100 bytes static

(6) Predicted Resource Availability: 2 Kbytes

(7) Data Availability Schedules Information: 90 bytes per DAS or 90 X 1500: 135 Kbytes per month for DASes corresponding to 1500 products (or 540 Kbytes per month for DASes corresponding to 6,000 products.)

(8) Planning Subsystem Parameters: 100 bytes

CONCLUSION: To support CERES (standard production) processing for about 1 month, Planning database requires about 40 - 42 Mbytes of disk space.

To support CERES processing for about 1 month (corresponding to standard and reprocessing production requests), Planning database requires about 164 Mbytes of disk space.

4.3   Database Access Rate for CERES Data Per Day -Epoch c

Given below are the major Planning functions performed in the Planning Subsystem on any given day and the number of accesses into the Planning database based on the number of PGE

activations per day of CERES production data (Epoch c). By database access is considered as a read/write operation of about 12 kbytes of data from/to the database.

(1) Production Request Processing:

Accepts about 1 Production Request per day.

Accesses to:

Production Request / PGE area of database:		1 (write)
Data Processing Request area	:	300 (write)
PGE Profile	:	10 (read)
Total	:	300 (approx.)

(2) Plan Creation: Generates 1 candidate plan to cover one month's worth of processing

(6000 DPRs)

Accesses into:

Data Processing Requests area	:	12,000 (read 6,000 DPRs)
Resource Availability information	:	1
Data Availability Schedule info	:	90 (read / write of DAS information)
PGE Profile	:	10 (read)
Plans area	:	4 (write)
Production Rules	:	1 (read)
Planning Parameters	:	1 (read)
Total	:	12,200 approx.

(3) Plan Activation: Activate 1 plan to cover one month's worth of processing (6000 DPRs)

Data Processing Requests area	:	30,000 (read / write of 6,000 DPRs plus read of additional 3,000 DPRs to cover access to other DPRs when checking activation rules, consolidation of old plan with new, etc.)
Data Availability Schedule area	:	90 (read / write of DAS information for active plan)
Plans area	:	16 (read / write of old and new plan)
Production Rules	:	1
Planning Parameters	:	1
Total	:	30,100 (approx.)

(4) Plan Feedback Update: Update plan with feedback from 200 DPRs per day

> Data Processing Requests area : 300 (to cover descendant DPRs that
>
>   may be canceled when a DPR fails)
>
>   X 2 (12kb for each access) X 2
>
>   (read/write) X 6 (for 6 statuses) =
>
>   7,200
>
> Plans   : 200 (write)
>
> Total   : 7,400 (approx.)

(5) Report Generation: On an average day, may have to generate a Data Processing Request status report, a Production Request Status report, a Planning Workload Report and a Planning Turnaround Time report for 200 DPRs.

> Accesses into Production Request and Data Processing Request area of database: 400

Approx. Total number of data base accesses into Planning database per day for CERES data processing: 50,400

5. Planning (CERES) Database Size Estimates (Epoch F)

5.1 Assumptions

The following has been estimated for CERES data using Epoch F average numbers:

1. Estimated for 1 month's worth of data.

2. 30 CERES PGEs supported

3. Only Standard Production requests considered.

   Average number of Standard Production Requests per month: 30.

   (To include reprocessing requests (multiply by 4): average number of production requests: 120 / month).

4. Average number of CERES Data Processing Request (corresponding to standard production requests) per day: 200 (based on approx. 200 activations per day).

   or 6000 Data Processing Requests per month (30 days).

   (Average number of CERES DPRs per day corresponding to production requests in general: 800 or 24,000 per month).

5. Average number of input files per CERES PGE: 200

6. Average number of output files per CERES PGE: 10

7. Assume about 150 - 200 Data Processing Requests per Production Request

8. Assume about 30 candidate plans and 30 active plans generated per month.

9. Assume about 10 process script templates associated with each PGE.

10. Assume database tracks only process script template filenames and actual templates stored on local disk storage.

11. Assume database environment used to generate each plan tracked is stored on local disk and the Planning database tracks only its location.

12. Assume 14 CERES resources are tracked by the Planning database.

13. Assume a fudge factor (of 2.69) not included when calculating database size.

14. Assume the Planning database tracks the following major data classes:

    - PGE Profile
    - Production Requests
    - Data Processing Requests
    - Plans

- Production Rules
- Predicted Resource Availability
- Data Availability Schedules
- Planning Subsystem Parameters

15. Estimation Rationale: Number of estimated accesses into the database is dependent on the number of processing activations per day per CERES instrument at LaRC.

5.2 Size Estimates of Planning (CERES) Database - Epoch F

One month's worth using EPOCH F statistics

(1) PGE Profile: 12 Kbytes per record

For 30 PGE records: 360 Kbytes (approx.) - Static value

(2) Production Requests: 1600 bytes

For 30 Standard Production Requests per month: 48 Kbytes

(For 120 Standard / Reprocessing Production Requests per month: 192 Kbytes)

(3) Data Processing Requests: 27 Kbytes per DPR

For 6000 Standard Data Processing Requests: (approx. 160 Mbytes) per month

(For 24,000 Data Processing Requests: (approx. 640 Mbytes) per month

(4) Plan: 192 Kbytes per plan

For 60 plans per month: 12 Mbytes

(5) Production Rules: 100 bytes static

(6) Predicted Resource Availability: 2 Kbytes

(7) Data Availability Schedules Information: 90 bytes per DAS or 90 X 24,000:

2.1 Mbytes per month for DASes corresponding to 24,000 products .

(8) Planning Subsystem Parameters: 100 bytes

CONCLUSION:

To support CERES processing for about 1 month (corresponding to standard and reprocessing production requests), Planning database requires about 655 Mbytes of disk space.

5.3 Database Access Rate for CERES Data Per Day - Epoch F

Given below are the major Planning functions performed in the Planning Subsystem on any given day and the number of accesses into the Planning database based on the number of PGE activations per day of CERES production data (Epoch F). By database access is considered as a read/write operation of about 12 kbytes of data from/to the database.

(1) Production Request Processing:

Accepts about 4 Production Request per day.

Accesses to:

Production Request / PGE area of database :		4 (write)
Data Processing Request area	:	1200 (write)
PGE Profile	:	30 (read)
Total	:	1200 approx.

(2) Plan Creation: Generates 1 candidate plan to cover one month's worth of processing

(24,000 DPRs)

Accesses into:

Data Processing Requests area	:	48,000 (read 24,000 DPRs)
Resource Availability information	:	1
Data Availability Schedule info	:	350 (read / write of DAS information)
PGE Profile	:	350 (read)
Plans area	:	1 (write)
Production Rules	:	1 (read)
Planning Parameters	:	1 (read)
Total	:	48,000 approx.

(3)   Plan Activation: Activate 1 plan to cover one month's worth of processing

(24,000 DPRs)

Data Processing Requests area	:	120,000 (read / write of 24,000 DPRs plus read of additional 12,000 DPRs to cover access to other DPRs when checking activation rules, consolidation of old plan with new, etc.)
Data Availability Schedule area	:	350 (read / write of DAS information for active plan)
Plans area	:	4 (read / write of old and new plans)
Production Rules	:	1
Planning Parameters	:	1
Total	:	120,000

(4) Plan Feedback Update: Update plan with feedback from 800 DPRs per day

Data Processing Requests area   : 12,00 (to cover descendant DPRs that

                                                  may be canceled when a DPR fails)

                                                  X 2 (12kb for each access) X 2

                                                  (read/write) X 6 (for 6 statuses) =

                                                  28,800

Plans   : 800 (write)

Total   : 30,000

(5) Report Generation: On an average day, may have to generate a Data Processing Request status report, a Production Request Status report, a Planning Workload Report and a Planning Turnaround Time report for 800 DPRs.

        Accesses into Production Request and Data Processing Request area of database: 1,600

Approx. Total number of data base accesses into Planning database per day for CERES data processing: 200,800

6. Planning (MODIS) Database Size Estimates (Epoch F)

6.1   Assumptions

The following has been estimated for MODIS data using Epoch F average numbers:

1. Estimated for 1 month's worth of data.

2. 34 MODIS PGEs supported

3. Only Standard Production requests considered.

    Average number of Standard Production Requests per month: 34.

    (To include reprocessing requests (multiply by 4): average number of production requests: 136 / month).

4.   Average number of MODIS Data Processing Request (corresponding to standard production requests) per day: 8,700 (based on approx. 8700 activations per day). or 261,000 (standard) Data Processing Requests per month (30 days). (Average number of MODIS DPRs per day corresponding to production requests in general: 34,800 or 1,000 K per month).

5. Average number of input files per MODIS PGE: 70

6. Average number of output files per MODIS PGE: 2

7. Assume about 7,000 Data Processing Requests per Production Request

8. Assume about 30 candidate plans and 30 active plans generated per month.

9. Assume about 10 process script templates associated with each PGE.

10. Assume database tracks only process script template filenames and actual templates stored on local disk storage.

11. Assume database environment used to generate each plan tracked is stored on local disk and the Planning database tracks only its location.

12. Assume 14 MODIS resources are tracked by the Planning database.

13. Assume a fudge factor (of 2.69) not included when calculating database size.

14. Assume the Planning database tracks the following major data classes:

    PGE Profile

    •   Production Requests

    •   Data Processing Requests

    •   Plans

    •   Production Rules

- Predicted Resource Availability
- Data Availability Schedules
- Planning Subsystem Parameters

15. Estimation Rationale: Number of estimated accesses into the database is dependent on the number of processing activations per day per MODIS instrument at GFSC.

6.2   Size Estimates of Planning (MODIS / GFSC) Database

One month's worth using EPOCH F statistics

(1) PGE Profile: 4.86 Kbytes per record

For 34 PGE records: 165 Kbytes (approx.) - Static value

(2) Production Requests: 56 Kbytes

For 34 Standard Production Requests per month: 1.9 Mbytes

(For 136 Standard / Reprocessing Production Requests per month: 7.6 Mbytes)

(3) Data Processing Requests: 8.5Kbytes per DPR

For 261,000 Standard Data Processing Requests: (approx. 2.2 Gbytes) per month

(For 1,044,000 Data Processing Requests: (approx. 8.9 Gbytes) per month

(4) Plan: 8 Mbytes per plan

For 60 plans per month: 480 Mbytes

(5) Production Rules: 100 bytes static

(6) Predicted Resource Availability: 2 Kbytes

(7) Data Availability Schedules Information: 90 bytes per DAS or 90 X 1.044,000: 90 Mbytes per month for DASes corresponding to 1.044,000 products .

(8) Planning Subsystem Parameters: 100 bytes

CONCLUSION:

To support MODIS processing for about 1 month (corresponding to standard and reprocessing production requests), Planning database requires about 9.5 Gbytes of disk space.

6.3   Database Access Rate for MODIS Data Per Day - EPOCH F

Given below are the major Planning functions performed in the Planning Subsystem on any given day and the number of accesses into the Planning database based on the number of PGE activations per day of MODIS production data (Epoch F). By database access is considered as a read/write operation of about 12 kbytes of data from/to the database.

(1) Production Request Processing:

Accepts about 4 Production Request per day.

Accesses to:

Production Request / PGE area of database :	18	(write)
Data Processing Request area :	20,000	(write)
PGE Profile :	14	(read)
Total :	20,000	approx.

(2) Plan Creation: Generates 1 candidate plan to cover one month's worth of processing

(1,000,000 DPRs)

Accesses into:

Data Processing Requests area :	750,000	(read 1,000,000 DPRs)
Resource Availability information :	1	
Data Availability Schedule info :	15,000	(read / write of DAS information)
PGE Profile :	14	(read)
Plans area :	670	(write)
Production Rules :	1	(read)
Planning Parameters :	1	(read)
Total :	766,000 approx.	

(3) Plan Activation: Activate 1 plan to cover one month's worth of processing (1,000,000 DPRs)

Data Processing Requests area: 1,875,000 (read / write of 1,000,000 DPRs

plus read of additional 500,000 DPRs to cover

access to other DPRs when checking activation

rules, consolidation of old plan with new, etc.)

Data Availability Schedule area :	1,500	(read / write of DAS information
		for active plan)
Plans area :	2680	(read / write of old and new plans)
Production Rules :	1	
Planning Parameters :	1	
Total :	1,880,000	

(4) Plan Feedback Update: Update plan with feedback from 34,800 DPRs per day Data Processing Requests area: 52,200 DPRs (to cover descendant DPRs that may be canceled when a DPR fails) X 2 (read/write) X 6 (for 6 statuses) = 450,000

Plans :	25,000
Total :	500,000 (approx.)

(5) Report Generation: On an average day, may have to generate a Data Processing Request status report, a Production Request Status report, a Planning Workload Report and a Planning Turnaround Time report for 34,800 DPRs.

Accesses into Production Request and Data Processing Request area of database: 25,000

Approx. Total number of data base accesses into Planning database per day for MODIS data processing: 3,190,000

Abbreviations and Acronyms

API	Application Programming Interface
DBA	Database Administration
DBMS	Database Management System
ECS	EOSDIS Core System
EOSDIS	Earth Observing System Data and Information System
ESDIS	Earth Science Data and Information System
PDPS	Planning and Processing System
PGS	Product Generation Subsystem
SDPS	Science Data Processing Segment